HP Doubles Up on Madison

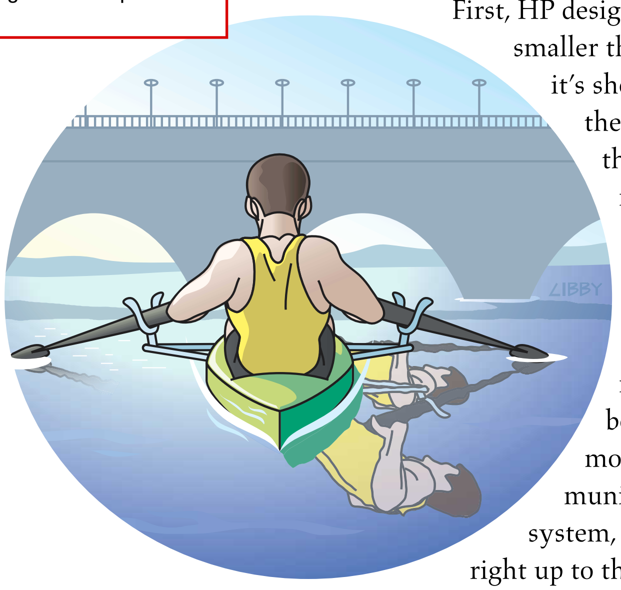
Quick Note

Gordon Haff

24 February 2003

As a partner in the development of Intel's IPF (Itanium Processor Family) architecture, Hewlett-Packard has always been among its most ardent supporters. Given that, today, HP's central strategy is to make IPF the common processor platform for all of its operating systems,[1] it's hard not to see HP as the system vendor at the forefront of Itanium-based computing. Consider just one statistic: in the third quarter of 2002, HP shipped 80% of the Linux running on IPF servers. Four times the rest of the market combined. That's a rout.

If you're IBM or Sun, how do you respond to this kind of dominance? Criticizing the chip's performance is one possibility, but that's a difficult tactic, considering that Itanium 2 has sprinted to a comfy spot among the frontrunners in the microprocessor performance race.[2] It's much simpler to knock Itanium 2 as a "commodity" processor and thereby imply that it will lag in top-end development because HP doesn't control the chip itself and will therefore not be able to innovate around the processor. But this argument is looking shaky, too.



First, HP designed an IPF chipset, the zx1, optimized for systems smaller than those targeted by Intel's own chipset efforts.[3] Now it's showing off a less conventional daughterboard design, the hp mx2 dual processor module (codenamed "Hondo") that it developed with the acknowledgement of Intel. The mx2 will provide HP IPF systems with a unique Itanium 2 performance option; they'll still be able to use off-the-shelf uniprocessor Itanium 2's, but now they'll also have the choice of plugging a dual-chip module into the exact same socket for better system performance on compute-intensive workloads. What's more, because a product like the mx2 requires design at the most intimate level of the system bus to coordinate communications between the processors and other parts of the system, it shows the willingness and ability of HP to engineer right up to the very edges of the processor core.

1. HP's "MultiOS strategy" for Itanium is primarily focused on Linux, Windows, and HP-UX—but OpenVMS and NonStop Kernel are also moving to Itanium. (Some major Tru64 technologies such as clustering will fold into HP-UX but Tru64 itself will remain Alpha-only.)
2. See Illuminata note "Itanium 2 Performance: *Wow!*" (August 2002).
3. See Illuminata note "HP's zx1: Chipset for Nimble McKinleys" (February 2002).

Making Two Be As One

The hp mx2 module's CPU horsepower comes from two second-generation Itanium 2 "Madison" processors[4] that run at a reduced clock speed to limit power consumption and heat production. They are powered by a single voltage regulator module (VRM) to save space. The other essential mx2 component is an HP-designed "Sherpa" bus converter and cache controller chip that makes the daughterboard play nice with the rest of the system. The mx2 module plugs into a standard Itanium 2 motherboard socket. It puts the same electrical load on the system bus as a standard uniprocessor Madison. It's the same size. It even has the same maximum power requirements. All this boring sameness means that the mx2 will be able to directly substitute for a regular single Madison processor in an HP system, boosting processing power without changing the physical layout of the machine—which isn't boring at all.

The Sherpa chip intermediates between the dual Madison processors on the mx2 and the external system bus into which the mx2 plugs, while effectively "fooling" both sides. To the processors, Sherpa makes itself look like a "central agent," effectively mimicking the external bus with which a processor would normally interact directly. To the external system bus, Sherpa looks just like a single Itanium 2 processor—thereby hiding the fact that there are really two complete processors on the module.[5] By looking like a CPU to the system bus and a system bus to the processors, Sherpa lets both sides see what they expect to see and thereby work properly.

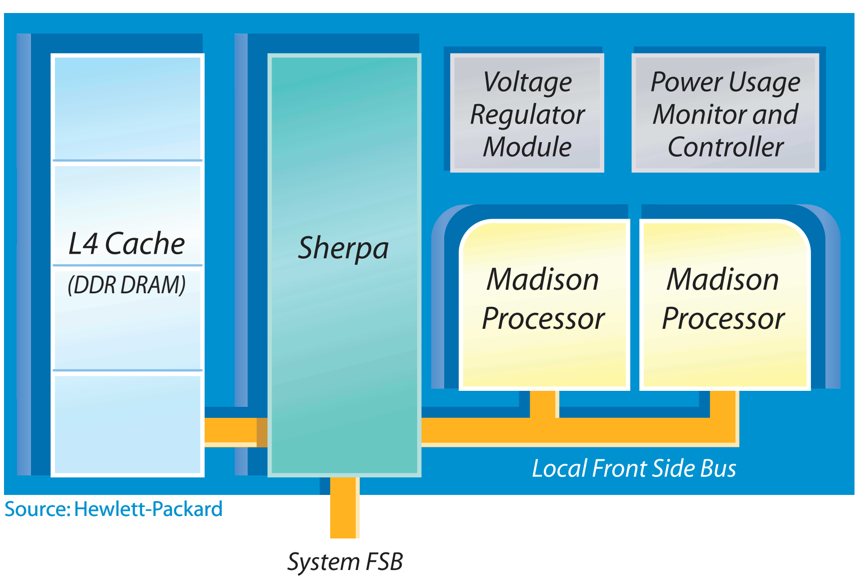
But Sherpa goes well beyond minimally substituting two processors for one. After all, there's little point in doubling up processors if system-level performance doesn't increase; and that often takes more than an additional CPU. Sherpa adds a large (32 MB) additional (Level 4) cache[6] to the already sizable caches on the Itanium processors themselves, in order to keep data close to the chip and improve performance even further. This DDR (double data rate) local cache memory lets the processors access recently-used data and instructions faster than fetching them from the relatively distant system

4. "Madison" is the second processor to carry the "Itanium 2" brand; the first was "McKinley."

5. At least from an electrical perspective. Software sees the mx2 as two full Madison CPUs.

6. This L4 cache complements the three levels of cache on the Itanium 2 processor, the largest of which—the L3—can be up to 6 MB. (Intel will also offer 3 MB and 4 MB variants.) The Madison processors in the mx2 will have standard caches, but HP has not yet decided which size(s) it will offer.

HP mx2 Module Block Diagram



Source: Hewlett-Packard

memory, without clogging up the paths to that memory. For example, in a four-socket (i.e., 8-CPU system using mx2 modules) system using HP's zx1 chipset, a main memory access takes about 62 clock cycles. But if the data's in the L4 cache, it only takes 12 cycles to retrieve—82 percent faster.[7] On a large system like Superdome—in which memory is more physically distributed and, therefore, can take longer to access—the latency difference between retrieving data from memory and a local cache will be even more pronounced.

The potential downside is that if the processor always checks the cache first for a chunk of data and only goes to memory if it's not in the cache, it can take longer to retrieve than if the cache weren't there. And application performance is often largely dependent on system latencies, not CPU counts.[8] HP's solution? Start requesting the data from memory at the same time it's requested from cache "just in case." By itself, this would be an imperfect approach because system memory would be bombarded by unnecessary, performance-draining requests for data being simultaneously retrieved from the local cache. However, Sherpa is designed in such a way that, if the data is found in the L4 cache, it has time to terminate the memory request before it's seen outside the mx2 module. This gives the mx2 the "keep data local" benefits of an additional level of cache without a latency penalty.

In some respects, the mx2 module presages the dual processor cores of Montecito—the next major IPF generation after Madison. But whereas Montecito won't arrive until about 2005 when a 90-nanometer process provides additional space on the chip for the second core, the mx2 will double-up the .13-micron process Madison processors that will ship later in 2003. (The mx2 is scheduled to ship in early 2004.)

7. Sherpa does add some memory access latency of its own—13 cycles—compared to a standard uniprocessor Madison connected directly to the system bus. However, this penalty applies only when data isn't in the L4 cache; it will be significant primarily on workloads in which large quantities of non-repeating data cause the caches to regularly flush and refill.
8. See Illuminata note "Latency Matters!" (Sept. 2002).

What of the Watts?

To appreciate some of the challenges involved in squeezing two Madison processors into a module that physically mimics a single chip, consider just the problem of power. How do you fit two Itanium 2 processors, to say nothing of the other module electronics, within the power envelope of a standard Madison—170 watts in all, 130 watts for the processor and 40 watts for its voltage regulator?

The first step is to run the chip's clock at less than the full 1.5 GHz of the top-speed Madison. This saves power,[9] but isn't by itself sufficient to do the job. A conventional worst-case power design would have had the mx2's dual processors running at ~800 MHz each, resulting in about the same number of total cycles split across two CPUs rather than one. But it's the rare workload indeed that would benefit from running across two half-speed CPUs rather than running on a single full-speed one. Instead, HP will run each Madison at a frequency intermediate between the worst-case scenario and its full rated speed.[10] The result thus increases performance, yet remains within its power budget.

The mx2 module is designed to run each processor so that it consumes about 65 percent of the max power that it could theoretically draw. HP figures this should handle all but truly oddball workloads that exercise both processors' execution units with worst-case data and instruction patterns for sustained periods. But the mx2 takes those pathological cases into account by dynamically monitoring the module's power use and, when need be, throttling the amount of processing the Madison does by switching it into "single issue mode."[11] Single-issue mode significantly erodes performance—typically slowing down the CPU by 75 percent—but it's an effective lever to keep the module within its design envelope in rare instances, while letting it run faster for more common workloads.

9. With CMOS semiconductors, higher frequencies mean more power consumed (and heat generated).
10. The exact GHz figure has yet to be publicly disclosed.
11. The Itanium 2 architecture can normally issue as many as six instructions simultaneously to its various execution units; typical applications can sustain about four issues per cycle.

Doubling Processor Counts, Boosting Performance

It's for those applications that make efficient use of additional processors rather than bottlenecking at disk, memory, or some other system component that the mx2 will shine. Fortunately there are many such application types; transaction processing, many types of high-performance computing, and business intelligence all typically scale quite well to the total processing power in the system rather than just the speed of an individual processor. For example, HP's simulations suggest that systems with mx2 modules will perform about 50 percent faster than those with single Madison CPUs of the same per-processor cache size when doing transaction processing. That's quite an increase for a drop-in replacement HP will offer as a processor-doubling option on its IPF servers[12] all the way up to its 64-way (128-way with mx2 modules) Superdome. Given some of the strong performance numbers already unveiled for IPF systems, the mx2 should help HP attain some truly impressive results.

However, as important as what the mx2 specifically does, is what it shows about how HP is developing in-house intellectual property that creeps ever closer to the very edge of the IPF processors themselves. Advanced system designs? Absolutely. Focused chipsets like the zx1? Sure. But developing a plug-compatible processor module that doubles the number of Itanium 2 processors in an HP system? That's a level of innovation at the deepest level that wasn't supposed to be possible with a "commodity" processor.

12. HP says that it will offer the mx2 module across a range of its server products, but stops short of committing to putting it in all its IPF servers.